MACHINE LEARNING

Q1 to Q11 have only one correct answer. Choose the correct option to answer your question.

1. Movie Recommendation systems are an example of:
   i) Classification
   ii) Clustering
   iii) Regression
   Options:
   a) 2 Only
   b) 1 and 2
   c) 1 and 3
   d) 2 and 3
   Ans-a)2 only

2. Sentiment Analysis is an example of:
   i) Regression
   ii) Classification
   iii) Clustering
   iv) Reinforcement
   Options:
   a) 1 Only
   b) 1 and 2
   c) 1 and 3
   d) 1, 2 and 4
   Ans-d)1,2 and 4

3. Can decision trees be used for performing clustering?
   a) True
   b) False
   Ans-a)True

4. Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:
   i) Capping and flooring of variables
   ii) Removal of outliers
   Options:
   a) 1 only
   b) 2 only
   c) 1 and 2
   d) None of the above
   Ans-a)1 only

5. What is the minimum no. of variables/ features required to perform clustering?
   a) 0
   b) 1
   c) 2
   d) 3
   Ans-b)1

6. For two runs of K-Mean clustering is it expected to get same clustering results?
   a) Yes
   b) No
   Ans-b)NO

MACHINE LEARNING

7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
   a) Yes
   b) No
   c) Can't say
   d) None of these
   Ans-a)Yes

8. Which of the following can act as possible termination conditions in K-Means?
   i) For a fixed number of iterations.
   ii) Assignment of observations to clusters does not change between iterations. Except for cases witha bad local minimum.
   iii) Centroids do not change between successive iterations.
   iv) Terminate when RSS falls below a threshold.
   Options:
   a) 1, 3 and 4
   b) 1, 2 and 3
   c) 1, 2 and 4
   d) All of the above
   Ans-d)All of the above

9. Which of the following algorithms is most sensitive to outliers?
   a) K-means clustering algorithm
   b) K-medians clustering algorithm
   c) K-modes clustering algorithm
   d) K-medoids clustering algorithm
   Ans-a)K-means clustering algorithm

10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):
   a. Creating different models for different cluster groups.
   b. Creating an input feature for cluster ids as an ordinal variable.
   c. Creating an input feature for cluster centroids as a continuous variable.
   d. Creating an input feature for cluster size as a continuousvariable.
   Options:
      i. 1 only
      ii. 2 only
      iii. 3 and 4
      iv. All of the above
   Ans-iv)All of the above

11. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?
   a) Proximity function used
   b) of data points used
   c) of variables used
   d) All of the above
   Ans-d)All of the above

MACHINE LEARNING

Q12 to Q14 are subjective answers type questions, Answers them in their own words briefly

12.Is K sensitive to outliers?

ANS: The K-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values. The algorithm seeks to minimise the squared Euclidean distances between the observation and the cluster centroid to which it belongs. However, the K-Means algorithm does not always produce the best results. It is susceptible to outliers. An outlier is a data point that differs from the rest of the data points.

13.Why is K means better?

ANS:Convergence is guaranteed with K means also centroids' positions can be warmed up. Adapts easily to new examples. Generalizes to different shapes and sizes of clusters, such as elliptical clusters. K means that it is easy to implement and adapts to new examples. With that we can also handle large data sets.

14.Is K means a deterministic algorithm?

ANS-The basic k-means clustering is based on a non-deterministic algorithm. This means that running the algorithm several times on the same data, could give different results. We can propose an improved, density-based version of K-Means that includes a novel and systematic method for choosing initial centroids.